

CalcZ Student Notes

Daniel Kaplan

2021-05-14

Contents

Status

This is the working version for Summer 2021.

Part I

Block 2: Differences and differentiation

Chapter 1

Functions and patterns

As you know, a function takes one or more inputs and returns a value as output. The functions we examine in CalcZ take *quantities* as inputs and return a *quantity* as an output. The algorithm that forms the body of the function describes arithmetic and other calculations that can turn the inputs into the output.

On the other hand, we can also use *tables* as functions. With a table, you specify the input, look up that input in one of the columns of the table which brings you to the right row. Then read out from that row the value in another column to be the output. The quantitative operation needed for table lookup is simple comparison. The floor/corridor/door metaphor describes table lookup as well as function evaluation.

In the previous block, we constructed functions to represent the patterns seen in data. In one example, we constructed a function $g(t) = A + Be^{-kt}$ to represent the temperature of water cooling in a mug as a function of time. In another example, we summarized the pattern of rising and falling tides.

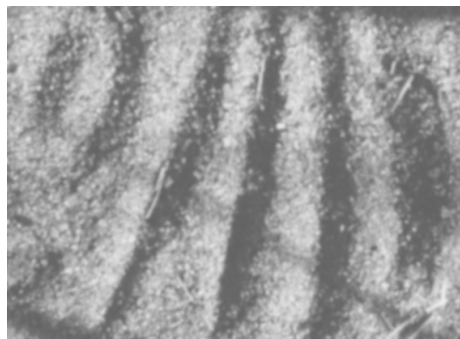
It's common sense that data is stored in tables. But we could easily represent any smooth mathematical function, such as our basic modeling functions, as a table look-up problem. Indeed, in the era before computers, many mathematical functions were used in exactly this manner: a printed table in which a person could search for a match to the input and retrieve a value for the output.

[Picture of some nice old table.]

In the computer era, we still routinely represent functions this way: data stored in computer files. For instance, an MP3 file is not much more than a sequence of numbers that record a complicated function of time: the air pressure variations of sound. Similarly, digital images record functions of x and y over a limited domain. Given x and y as input, you can look up the output by going to the right pixel.

We humans of course, don't perceive the numerical output of either sound or image functions: we **hear** a sound and we **see** an image. We've got special biological equipment for this!

Consider the image in Figure @ref{fig:sand-1}. It is a picture of some indentations in a small area of sand, about two inches wide in the middle of a hiking trail. The dots are individual grains of sand.



Can you see three almost parallel furrows? How about the small crater in the upper left?

You can see the individual grains of sand because they contrast sharply with their neighbors. Or, more precisely, you can see a few dozen grains **because** they contrast sharply with their neighbors.

You can think of the surface of the sand as a function of x and y . It's lower in some places and higher in others. But, in fact, you can't see the height of an individual point in the photograph. In the right light, you wouldn't notice the furrows at all. But the way the picture is lighted, raking sunlight from the left, the surface is translated into broad regions of brightness and shadow. In the light regions, the surface slants toward the sun. In the shadows, the surface slants away from the sun.

What you're mainly seeing in the photo is the ***slant*** or ***slope*** of the surface. The light has transformed *elevation* as a function of x and y into *slant* as a function of x and y and then encoded the slant as brightness, in much the same way the background of a contour plot encodes the output of the function as a color.

The moral here is that sometimes the data in a function is not in the right form for us to extract useful information. But by transforming that data to represent contrast or difference or slope, the information can be revealed.

This Block is about transforming functions to show difference and slope. Such transformation, accomplished by mathematics rather than the raking light of the sun, can take a pattern that we're presented with and turn it into another pattern that can tell us what we want to know.

1.1 Average rate of change

1. **[Deriv-2a]** *Identify the “average rate of change” of a function over an interval as the overall change in output divided by the change in input.*

As you know, one way to calculate the slope of a straight-line function $f()$ is to pick two different values for the input: call the larger of them x_1 and the smaller x_0 . Think of these input values as the endpoints of an **interval** or **neighborhood** of the domain. The length of this domain is $x_1 - x_0$.

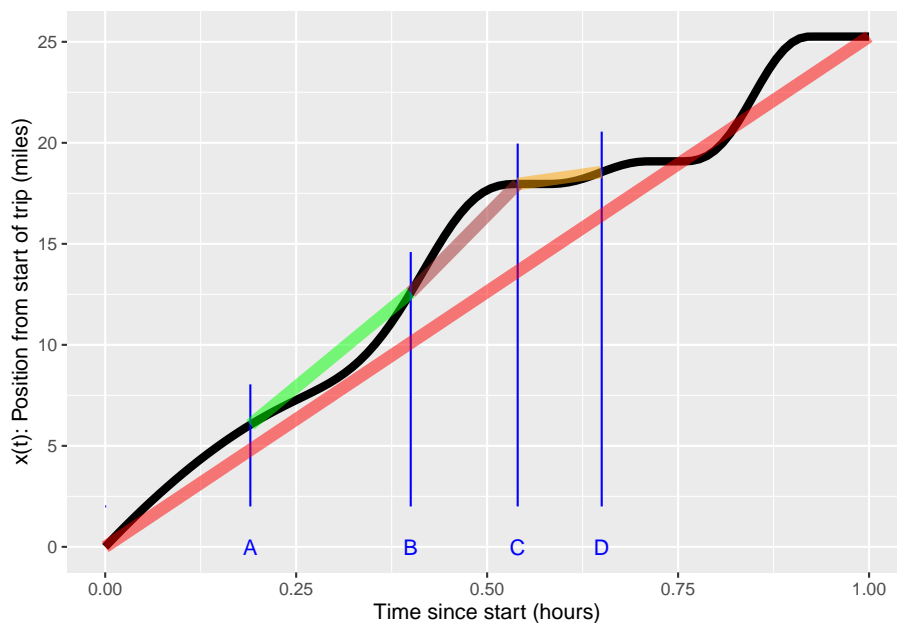
Evaluate the function at those endpoints and calculate the slope simply as the difference in output divided by the length of the interval:

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

“Slope” is a natural metaphor when thinking of a function as a graph. But a more general way to describe the concept is the **rate of change** of the output with respect to the input. The change in the output from one end of the interval is $f(x_1) - f(x_0)$, the change in the input is $x_1 - x_0$. If the input is time (in hours), and the output is the position of a car (in miles), then the rate of change is *miles-per-hour*: the car’s velocity.

For a straight-line function—think of a car driving at constant speed on a highway—it doesn’t matter what you choose for x_1 and x_0 (so long as they are not identical). But for other functions, the choice does matter.

Imagine a graph of the position of a car along a road as in Figure @ref{fig:stop-and-go}. Over the course of an hour, the car travelled about 25 miles. In other words, the **average** speed is 25 miles/hour: the *slope* of the red line segment. Given the traffic, sometimes the car was stopped (time C), sometimes crawling (time D) and sometimes much faster than average (time B).



During the interval from B to C, the car was travelling relatively fast. The slope of the brown segment connecting the position at times B and C is the *average* rate of change between times B and C. It's easy to see that the average rate of change from B to C is larger than the overall average from $t = 0$ to $t = 1$. Calculating that slope is a matter of evaluating the position at the endpoints and dividing by the length of the interval.

What is the average rate of change in the car's position during the interval $t_B = 0.40$ to $t_C = 0.54$?

The length of the interval is $t_C - t_B = 0.54 - 0.40 = 0.14$.

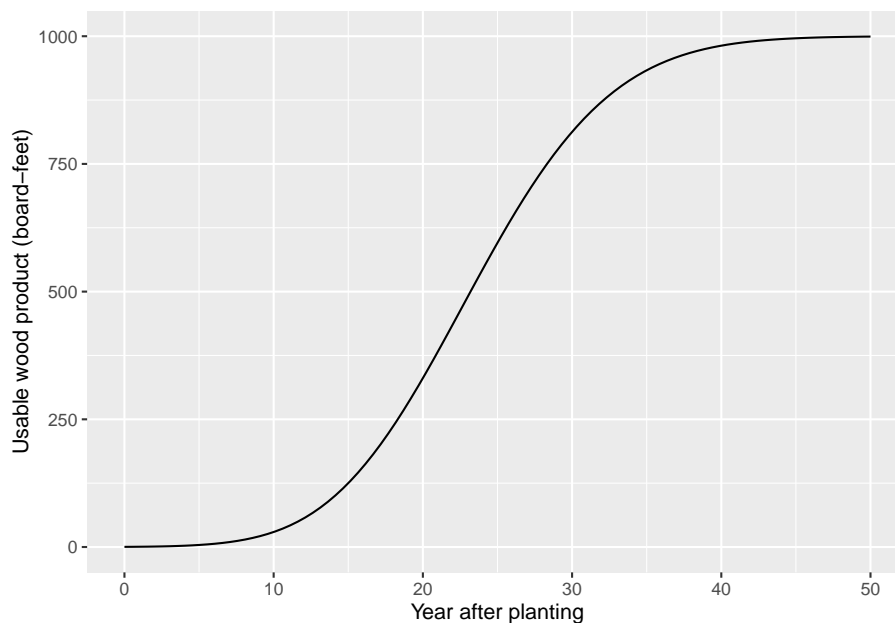
Evaluating the function gives $x(t_C) = 18$ and $x(t_B) = 12.6$.

Rise is $x(t_C) - x(t_B) = 18 - 12.6 = 5.4$.

Run is $t_C - t_B = 0.54 - 0.40 = 0.14$.

The average rate of change during the interval is $5.4/0.14 = 38.6$ miles/hour.

The graph shows a simplified model of the amount of usable wood that can be harvested from a typical tree in a managed forest of Ponderosa Pine. (You can see some actual forestry research models [here](#).)



You are writing a business plan for a proposed pine forest. Among other things, you have to forecast the revenue that will be generated and when you will have salable product.

They say that “time is money.” Every year you wait before harvest is another year that you don’t have the money. On the other hand, every year that you wait means more wood at the end. How to decide when to harvest?

The tree continues to grow until year 50, when it seems to have reached an equilibrium: perhaps growth goes to zero, or rot balances what growth there is. There’s no point waiting until after year 50.

At year 25, the tree is growing as fast as it ever will. You’ll get about 600 board-feet of lumber. Should you harvest at year 25? No! That the tree is growing so fast means that you will have a lot more wood at year 26, 27, and so on. The time to harvest is when the growth is getting smaller, so that it’s not worth waiting an extra year.

The quantity of interest is the average rate of growth from seedling to harvest. Harvesting at year 25 will give a total change of 600 board feet over 25 years, giving an average rate of change of $600 \div 25 = 24$ board-feet-per-year. But if you wait until year 35, you’ll have about 900 board feet, giving an average rate of change of $900 \div 35 = 25.7$ board-feet-per-year.

We’ve been presenting the average rate of change as a **number**:

$$\frac{f(t_B) - f(t_A)}{t_B - t_A}$$

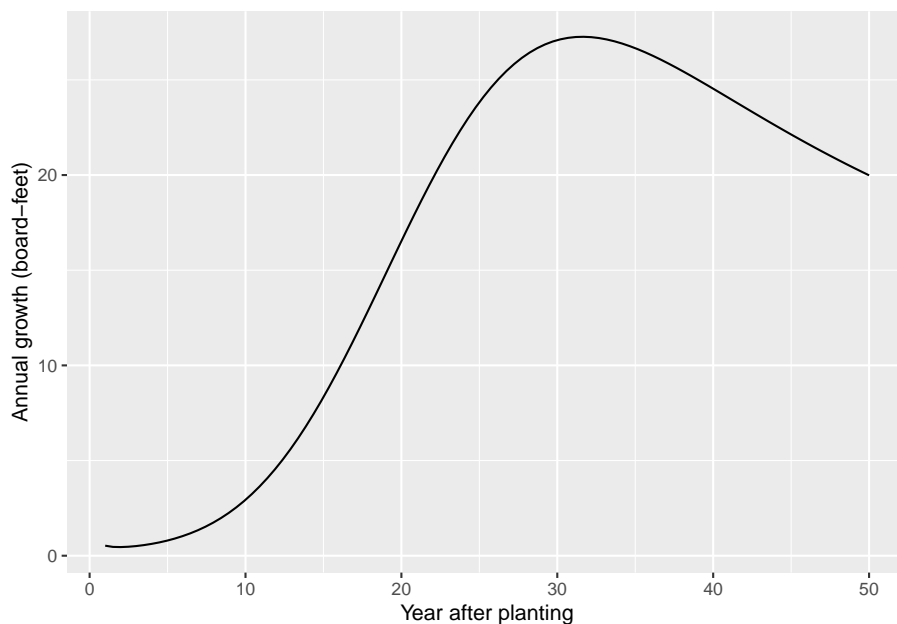
But it's helpful to think of it as a **function** that takes two inputs, which we can call t_A and t_B :

$$\text{ave_rate_of_change}(t_A, t_B) \equiv \frac{f(t_B) - f(t_A)}{t_B - t_A}$$

This is a very subtle maneuver. The formula is exactly the same, but by writing the quantities t_A and t_B as inputs (by putting them in parentheses on the left side), we turn the formula into a function.

Sometimes the average-rate-of-change function presents the information in a clear way.

Back to the forest ... Here's a graph of the average-rate-of change function. The function has two inputs, but botany dictates that we can only be growing wood from the time the seedling is planted. So we'll set $t_A = 0$ and look at the corresponding *slice* of the function:



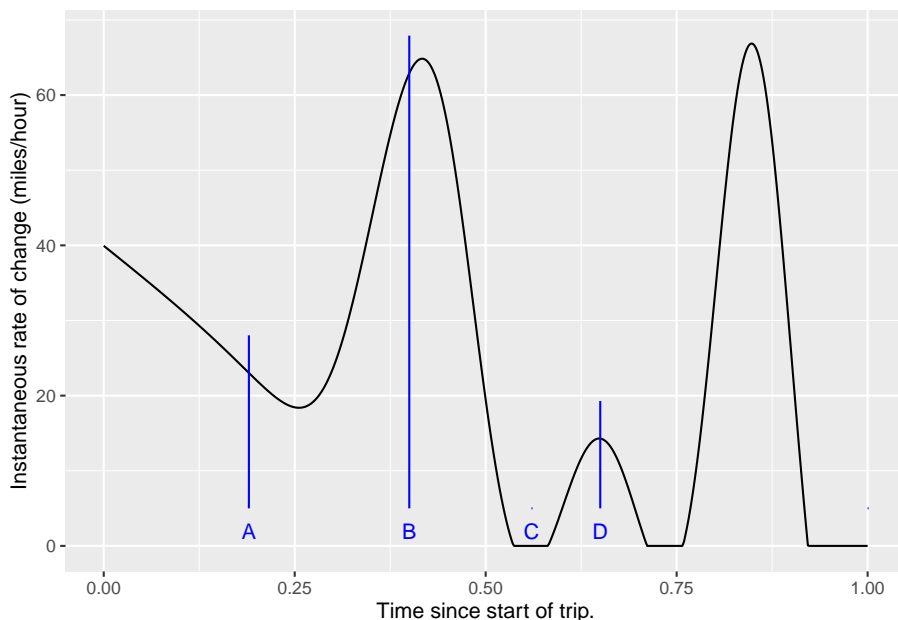
The graph makes it clear that the maximum average growth from planting to harvest will occur at about year 32.

1.2 Instantaneous rate of change

1. [Deriv-2b] Distinguish the “average rate of change” from the “instantaneous rate of change.”

The car’s speedometer shows the speed at each moment—or *instant*—of the trip. As you can see in Figure @ref{fig:stop-and-go}, the speed varies and is sometimes less than the average speed, sometimes greater, and occasionally equal to the average speed over the trip. The general term for the kind of quantity presented by the speedometer is the *instantaneous rate of change* of the position function with respect to the input to that function.

Figure @ref{fig:instant-speed} shows the instantaneous rate of change of position with respect to time.



The two graphs in Figures @ref{fig:stop-and-go} and @ref{fig:instant-speed} show exactly the same car trip. The presentation of the data in the different graphs makes it easy to see some things and hard to see others. For instance, figuring out when the car is at a stand-still is harder in the position-vs-time graph than in the speed-vs-time graph. This is very much in the spirit of the sand-furrows example at the start of this chapter: it’s much easier to perceive the furrows because the lighting highlights areas sloping toward the sun as bright and areas sloping away from the sun as dark. In Figure @ref{fig:instant-speed} we’re not using light-and-dark for the display. Instead, we’re showing the instantaneous speed as the vertical position

Recall that the interval between t_B and t_C had an *average rate of change* of about 39 miles-per-hour. Looking at the *instantaneous rate of change* tells the story differently: at time t_B the car was accelerating to about 60 miles-per-hour. Then it gradually slowed, coming to a stop just before time t_C .

Figure @ref{fig:stop-and-go} shows the function $\text{position}(t)$. Figure @ref{fig:instant-speed} shows a different function, $\text{speed}(t)$. Although

the two functions are different, they are intimately related: $\text{speed}(t)$ is the *instantaneous rate of change* of $\text{position}(t)$.

Two central operations in calculus are:

1. Given a function $f(t)$, find the function $g(t)$ giving the instantaneous rate of change of $f()$. This process of deriving $g(t)$ from $f(t)$ is called *differentiation*.
2. Given a function $g(t)$, find the $f(t)$ of which $g(t)$ is the instantaneous rate of change. This process of finding $f()$ given $g()$ is called *anti-differentiation*.

YOU WERE WORKING HERE.

Sometimes, working with the average-rate-of-growth function directly tells you what you need. Other times, you need to extend your model.

Between year 30 and 32, there is hardly any change in the value of the average-rate-of-change function. It's increasing a little, but is it really worthwhile to wait. One argument is that at year 29 you already have a valuable resource: wood that could be money in the bank. If the money were in the bank, you could invest it and earn more money *and* at the same time get a new seedling in the ground to start its growth. You're doing two things at once. Efficient!

To know what is the best year for harvest from this point of view, you want to calculate the effective "interest rate" on the present amount of wood that you earn in the form of new wood. That interest rate is the

Chapter 2

Instantaneous rate of change

2.1 From average to instant

YOU HAVE KNOWN HOW TO CALCULATE AN AVERAGE RATE OF CHANGE for a long time. But possibly you don't yet fully know what the **use** of such a calculation is other than to help you pass a math test. In this section, we're going to show how to calculate the *instantaneous rate of change* function and show some of the many things this is useful for.

To orient you to what's going to happen ...

1. We're going to set up the familiar average-rate-of-change calculation in a new framework.
2. Using that framework, you'll be able to calculate an average-rate-of-change *function*.
3. Then, by turning a knob on the calculation in (2), you'll be able to calculate the *instantaneous* rate of change function. By the way, the knob is called h .

In short, this section will introduce you to the operation called *differentiation*.

2.2 The h framework

Calculating an average rate of change involved specifying an interval of the function domain: the left side t_A and the right side t_B . Let's change this notation a little:

- Left endpoint of domain: t_0
- Right endpoint of domain: $t_0 + h$

There is nothing fundamentally new here. The interval is still specified by two numbers, but now they are called t_0 and h . The width of the interval is just h , which is a little easier to write than $t_B - t_A$, but is still just the width.

The **average rate of change** in the new notation is

$$\frac{f(t_0 + h) - f(t_0)}{h} \quad \text{rather than} \quad \frac{f(t_B) - f(t_A)}{t_B - t_A}$$

Why are you using t instead of x ?

Remember, the *name* of an input can be anything at all so long as you use it consistently on the left side of \equiv and on the right side. All of these are the same definition:

$$g(t) \equiv e^{kt}, \quad g(x) \equiv e^{kx}, \quad g(y) \equiv e^{ky}, \quad g(\text{altitude}) \equiv e^{k \text{altitude}}$$

When we work with functions of two or more variables, it will be essential to give easily distinguished names to the inputs. We're trying to get you in the habit of paying attention to the names of inputs and break a *bad habit* from high-school math of calling the input x and the output y .

Now we are going to make a small, subtle, and important change. Instead of thinking of the average rate of change *at a specific interval* t_0 to $t_0 + h$, we're going to write an average-rate-of-change **function**. And since we're in the habit of giving *names* to functions, we'll be careful to name average-rate-of-change functions in a way that makes explicit the connection to the function $f(t)$ whose rate of change is being calculated.

Our convention will be to use two additional components in the names of average-rate-of-change functions. The first component is to lead with a “D” in a caligraphic font: D . The second component will be the name of the input whose interval is being used in calculating the rate of change. So ...

$$D_t f(t) \equiv \frac{f(t + h) - f(t)}{h}$$

You can pronounce D_t as “difference with respect to t .” But remember that the function name is the whole deal, $D_t f()$, which is to be read as “difference with respect to t of $f()$.” Admittedly a long name. Don't be surprised later when we start using *nicknames* like Liz (for Elizabeth) or Bill (for William). For instance, later we'll save ink and breath by using the nickname $\dot{f}()$ instead of $D_t f()$.

FOLLOWING NEEDS TO BE SORTED OUT ...

From the way we've defined $D_t f(t)$, it's reasonable to assume that h is a **parameter**: a symbol naming a numerical value that has to be specified before $D_t f(t)$ can be evaluated at a specific t .

Find an average-rate-of-change function with respect to x of $g(x) \equiv x^2$.

$$D_x g(x) = \frac{(x+h)^2 - x^2}{h} = \frac{x^2 + 2xh + h^2 - x^2}{h} = \frac{2xh + h^2}{h} = 2x + h$$

I like to think of h as a kind of *tire iron*, a small tool used to stretch the bead of a bicycle tire in order to pull it over the wheel rim.



Figure 2.1: A tire iron in use

Once the tire iron has done its job, it's removed and you would never know that it was ever there (except that the tire is now successfully mounted on the wheel).

Find an average-rate-of-change function with respect to x of $g(x) \equiv x^2$, but remove the tire iron h when you're done.

$$D_x g(x) = \frac{(x+h)^2 - x^2}{h} \text{ stretch } x \text{ a bit} = \frac{x^2 + 2xh + h^2 - x^2}{h} \text{ pull over rim} = \frac{2xh + h^2}{h} \text{ still pulling } \dots = 2x + h \text{ su}$$

Now remove the tire iron to get

$$D_x g(x) = 2x$$

But this is calculus, not bicycle mechanics. How do we know that removing the tire iron isn't damaging the mathematical wheel? Historically, this has been a serious debate, resolved only with great difficulty more than a century after calculus started being used successfully.

Still in the spirit of having fun, let's try a more serious metaphor... imagining h is actually a central character in a calculus play. The character h is in the

middle of the story but *never appears in the play*, like the missing character Godot in the famous play *Waiting for Godot*.

We said that h in the finite rate of change function $D_x g()$, so long as h is small, plays both a central role and has hardly any effect. An economizing director re-writes the play to take h out of it, setting $h = 0$: a non-speaking, offstage role.

We’ve already seen using legitimate algebra that

$$D_x g(x) = 2x + h$$

Re-writing by replacing h with 0 streamlines the play, turning $\Delta g()$ from a dialog involving both x and h into a monologue with h absent:

$$D_x g(x) = 2x$$

. Simple.

And yet ... the director gets a letter from the Bit Players Union.

We observe that you have eliminated the role of h in the final production version of $D_x g(x)$. This is a violation of Union regulations. Recall that the rate-of-change function $D_x g(x)$ is defined as a ratio:

$$D_x g(x) \equiv \frac{g(x+h) - g(x)}{h}$$

*Although the name h does not need to appear in the argument list of $D_x g(x)$, eliminating h entirely by replacing her with zero is a **division by zero error** forbidden by Article 3.16§B¶2 of the Unified Laws of Arithmetic. We ask that you comply with this Article by re-instating the role of h in all evaluations of $Dg(x)$.*

Reading this, the director calls her lawyer. Is there a loophole for removing h without breaking the mathematical prohibition on dividing by zero?

2.3 The derivative operator

Let’s put aside for the moment the issue of the disappearing h . Historically, such “putting aside” has incurred a great deal of criticism. In 1734, famous philosopher George Berkeley (1685-1753) published a long-titled book: *The Analyst: A Discourse Addressed to an Infidel Mathematician: Wherein It Is Examined Whether the Object, Principles, and Inferences of the Modern Analysis Are More Distinctly Conceived, or More Evidently Deduced, Than Religious Mysteries and Points of Faith*. In *The Analyst*, Berkeley took issue with the arguments of that time that it is legitimate to divide by h when, ultimately, h will be replaced by zero. Calling h an “evanescent increment,” he asked,

“And what are these same evanescent Increments? They are neither finite Quantities nor Quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?”

Interesting, Berkeley believed that the ghost of h yielded correct results. His objection was that the framers of calculus had made two, canceling errors.

“[B]y virtue of a two fold mistake you arrive, though not at science, yet truth.”

Berkeley was saying that calculus had not yet been put on a solid logical foundation. It was more than a century after Berkeley’s death that this work was accomplished. Once accomplished, the results that had been claimed true all along were confirmed.

I propose that we start with the results, which is what everyone *uses*. Later, we can introduce the new concepts on which the new logic was based. The names of the concepts—continuity, smoothness, singularity—are widely and effectively used in talking about functions, but the names, like many words in English, can be put to good and accurate use without memorizing the precise definitions.

On to the results ...

1. Every “smooth” function $h(x)$ has a corresponding function that is its derivative $\partial_x h(x)$.
2. Finding the algorithm for $\partial_x h(x)$, called ***differentiating*** $h()$, is *automatic* in the sense that it can be done by computer without human intervention or judgment.
3. When there is an algebraic formula for $h(x)$, the computer can find the algebraic formula for $\partial_x h(x)$. This is called ***symbolic differentiation***.
4. When there is no algebraic formula for $h()$, the computer can use ***numerical differentiation*** which involves a non-zero h . The results of numerical differentiation on functions with formulas can differ subtly from the results from symbolic differentiation. Usually the difference is too small to notice, but in extreme cases it is not. It’s important for the user of numerical differentiation to know how to identify extreme cases and deal with them.

YOU WERE HERE.

Often, we use a function so briefly that it’s not worth naming. For instance, rather than saying

“Consider the function $h(x) = \sin(2\pi x/P)$ and the corresponding average rate of change function $D_x h(x)$...”

We’ll write: $D_x[\sin(2\pi x/P)]$

2.4 Evanescent h

The word “evanescent” means, “lasting for only a short time, then disappearing quickly and being forgotten.” (Source) Another dictionary definition is “tending to vanish like vapor.” (Source) I’ve used metaphors for h : tire irons, dramatic characters who never appear or speak. Here’s another one: the solvent in ink or paint. Ink and paint are liquid yet become solid when brushed on a surface. This can happen because the solid pigments are suspended in liquid. Exposed to the air, the liquid vaporizes and is forgotten. The solid pigment remains on the surface.

We can easily evaluate an **average rate of change function**, for instance $D_t g(t)$ by evaluating $g(t)$ at two different inputs and dividing by the distance between the inputs:

$$D_t g(t) = \frac{g(t+h) - g(t)}{h}$$

So if you know $g()$, you know $D_t g(t)$. By making h very small, the “average” starts to look like “instantaneous.” Over a small interval of length h , every smooth function looks like a straight line. For a straight line, the average rate of change is exactly the same as the instantaneous rate of change.

We’d like to take h out of the picture, imagining that it is zero but not dealing with “divide by zero” if we can avoid it. To signify this idea of neglecting h , we’ll use a new notation. Rather than $D_t g(t)$ which has that pesky h in it, we’ll write $\partial_t g(t)$, using the lower-case ∂ . The function $\partial_t g(t)$ can properly be called the **instantaneous rate of change** of $g(t)$. But everyone calls it the **derivative of $g(t)$ with respect to t** . Many people leave out the “with respect to” part, which I suppose is fine for a function like $g(t)$ which has only one input. But functions constructed in modeling generally have multiple inputs, so it’s a bad habit to neglect to identify which one the derivative is “with respect to.” We’ll always sneak the “with respect to” by putting the little subscript on ∂_t .

We lose something by letting h vaporize. There is an easy formula for $D_t g(t)$, but without h we can’t write down such a formula for the derivative. Mathematicians like to bridge the gap between D and ∂ , that is between average and instantaneous rates of change, by writing the formula for the derivative with a little “caution” sign out front.

$$\partial_t g(t) = \lim_{h \rightarrow 0} \frac{g(t+h) - g(t)}{h}$$

The caution sign is mysterious to newcomers because it is saying something absurd: “Remember to forget h .”

There are people who do actual calculations with $\lim_{h \rightarrow 0}$ but nobody needs to any more unless they invent a brand-new function that can’t be written with existing functions. If you do happen to invent such a brand-new function, please remember that you should also invent the derivative of that function.

Fortunately for us, the inventors of the past have done the work for us. Here's a list of what they found for the mathematical functions that underlie our basic modeling functions:

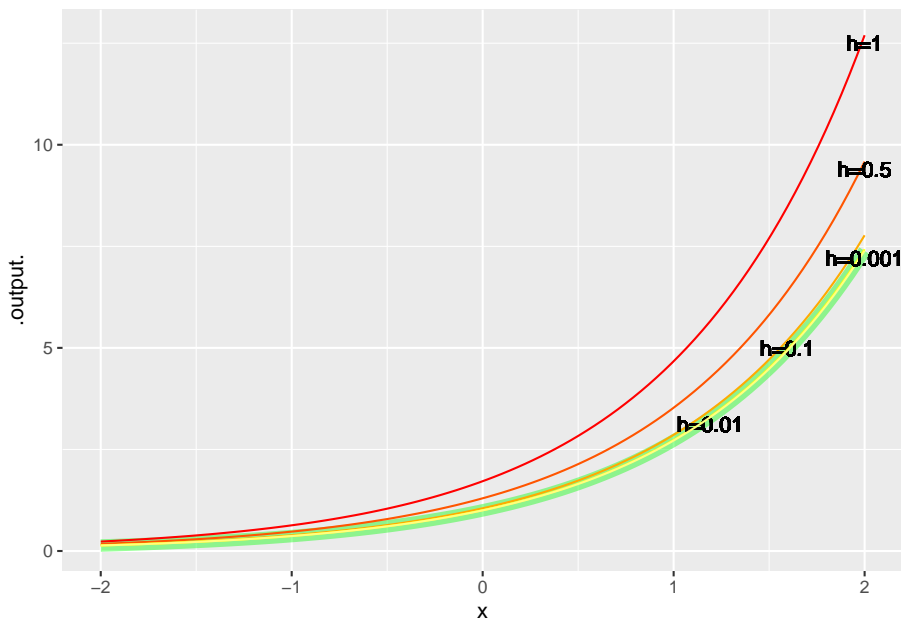
- $\partial_x e^x = e^x$
- $\partial_x x^0 = 0$
- $\partial_x x^p = p x^{p-1}$. unless $p = 0$.
- $\partial_x \sin(x) = \cos(x)$
- $\partial_x \cos(x) = -\sin(x)$
- $\partial_x \text{sigmoid}(x) = \text{hump}(x)$

We're going to leave $\partial_x \text{hump}(x)$ alone for now until we have a more complete understanding of derivatives. But other than that, notice that the derivative of every one of these mathematical functions—exponential, power-law, sinusoid—is also in the set of the basic functions.

How can we know that these formulas are right? Let's start by comparing them to the functions computed from the D_x formula when h is very small.

Consider $\partial_x e^x$, which is claimed to be e^x . Figure @ref{fig:exp-h} plots e^x (in green) as well as a function $\text{ARC}(x, h)$ (**A**verage **R**ate of **C**hange) that takes both x and h as inputs:

$$\text{ARC}(x, h) \equiv D_x e^x = \frac{e^{x+h} - e^x}{h} = e^x \frac{e^h - 1}{h}$$



You can see at a glance that $\text{ARC}(x, h)$ gets closer and closer to the claimed derivative, e^x , as h gets smaller.

A better view of things comes from recognizing that

$$\text{ARC}(x, h) = e^x \left[\frac{e^h - 1}{h} \right]$$

Let's look at the quantity in brackets as h gets small. If that goes to 1 for small h , then $\text{ARC}(x, h)$ goes to e^x .

	$\frac{e^h - 1}{h}$
1	1.71828182845905
0.1	1.05170918075648
0.01	1.00501670841679
0.001	1.00050016670838
0.0001	1.00005000166714
0.00001	1.00000500000696
0.000001	1.00000049996218
0.0000001	1.00000004943368
0.00000001	0.999999993922529
0.000000001	1.00000008274037

NEXT STEP. The linear expansion ...

$$f(x + h) \approx f(x) + h \partial_x f(x)$$

$$D f(g(x)) = f(g(x+h)) - f(g(x)) = f$$

How we When we are interested in an *instantaneous rate of change function*,

YOU GOT HERE.

2.5 The Δ operator

CHANGE THIS TO D and introduce it once you have gotten rid of h .

Up until now, we used only functions that take numerical quantities as inputs and produce a numerical quantity as an output. It's time now to broaden our perspective a bit.

Imagine a function named $\Delta()$ defined like this:

$$\Delta(f) \equiv \frac{f(x+h) - f(x)}{h}$$